A NAC Gene Regulating Senescence Improves Grain Protein, Zinc, and Iron Content in Wheat

Cristobal Uauy, *et al.*

Science **314**, 1298 (2006);
DOI: 10.1126/science.1133649

multilayer endosperm. However, most of the iron in rice seed, for example, is associated with the embryo and the aleurone layer, not the endosperm, suggesting that VIT1-mediated iron storage in the embryo may play the same role in developing endospermic plants as that described here for *Arabidopsis*. Furthermore, unlike other Fe transporters characterized to date such as IRT1, which can transport Cd as well as Fe (*21*), VIT1 does not appear to transport Cd. Cd levels in seeds from lines overexpressing VIT1 were low (< 0.1 part per million), with no significant difference compared to wild-type seeds ($P < 0.05$). Therefore, any potential biotechnological applications of VIT1 will not have to consider unwanted accumulation of this toxic heavy metal.

Our study demonstrates the power of combining mutant analysis with a technique that can both image and determine the elemental composition of living plant material. Although 2D imaging with x-ray fluorescence has been used before to image the distribution of metals in plant tissues (*22*, *23*), including *Arabidopsis* seed (*24*), our ability to render 3D images at high resolution allowed us to determine that Fe was associated with the provascular system throughout the seed and should prompt more studies on spatial distribution of metals in biological samples. Our study also highlights the role of the vacuole in seed iron storage and suggests that the vacuole offers another avenue for increasing the iron content of plant-based diets.

References and Notes

1. C. Curie, J.-F. Briat, *Annu. Rev. Plant Biol.* **54**, 183 (2003).
2. F. Goto, T. Yoshihara, N. Shigemoto, S. Toki, F. Takaiwa, *Nat. Biotechnol.* **17**, 282 (1999).
3. P. Lucca, R. Hurrell, I. Potrykus, *J. Am. Coll. Nutr.* **21**, 184S (2002).
4. L.-Q. Qu, T. Yoshihara, A. Ooyama, F. Goto, I. Takaiwa, *Planta* **222**, 225 (2005).
5. A. Pich, R. Manteuffel, S. Hillmer, G. Scholz, W. Schmidt, *Planta* **213**, 967 (2001).
6. F. Raguzzi, E. Lesuisse, R. R. Crichton, *FEBS Lett.* **231**, 253 (1988).
7. H.-P. Bode, M. Dumschat, S. Garotti, G. F. Fuhrmann, *Eur. J. Biochem.* **228**, 337 (1995).
8. L. Li, O. S. Chen, D. M. Ward, J. Kaplan, *J. Biol. Chem.* **276**, 29515 (2001).
9. O. S. Chen, J. Kaplan, *J. Biol. Chem.* **275**, 7626 (2000).
10. M. Schmid *et al.*, *Nat. Genet.* **37**, 501 (2005).
11. J. L. Bowman, S. G. Mansfield, in *Arabidopsis: An Atlas of Morphology and Development*, J. Bowman, Ed. (Springer-Verlag, New York, 1994), pp. 351–361.
12. D. Eide, M. Broderius, J. Fett, M. L. Guerinot, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5624 (1996).
13. N. J. Robinson, C. M. Proctor, E. L. Connolly, M. L. Guerinot, *Nature* **397**, 694 (1999).
14. E. E. Rogers, M. L. Guerinot, *Plant Cell* **14**, 1787 (2002).
15. E. P. Colangelo, M. L. Guerinot, *Plant Cell* **16**, 3400 (2004).
16. J. N. Lott, M. M. West, *Can. J. Bot.* **79**, 1292 (2001).
17. M. S. Otegui, R. Capp, L. A. Staehelin, *Plant Cell* **14**, 1311 (2002).
18. S. R. Sutton *et al.*, in *Applications of Synchrotron Radiation in Low-Temperature Geochemistry and Environmental Science*, P. A. Fenter, M. L. Rivers, N. C. Sturchio, S. R. Sutton, Eds. (Mineralogical Society of America, Washington, DC, 2002), vol. 49, pp. 429–483.
19. W. H. Stuppy, J. A. Maisano, M. W. Colbert, P. J. Rudall, T. B. Rowe, *Trends Plant Sci.* **8**, 2 (2003).
20. V. Lanquar *et al.*, *EMBO J.* **24**, 4041 (2005).
21. G. Vert *et al.*, *Plant Cell* **14**, 1223 (2002).
22. T. Punshon, A. Lanzirotti, S. Harper, P. M. Bertsch, J. Burger, *J. Environ. Qual.* **34**, 1165 (2005).
23. D. H. McNear Jr. *et al.*, *Environ. Sci. Technol.* **39**, 2210 (2005).
24. L. W. Young, N. D. Westcott, K. Attenkofer, M. J. T. Reaney, *J. Synchrotron Radiat.* **13**, 304 (2006).
25. S. Kumar, K. Tamura, I. Jacobsen, M. Nei, *MEGA2: Molecular Evolutionary Genetics Analysis Version 2.0.* (Pennsylvania and Arizona State Universities, University Park, PA, and Tempe, AZ, 2000).
26. J. S. Busse, R. F. Evert, *Int. J. Plant Sci.* **160**, 1 (1999).

27. We thank F. Ausubel, C. R. McClung, and D. Salt for critically reading the manuscript; B. Lahner and D. Salt for ICP-MS analysis; and A. Lavanway for help with microscopy. This work was supported by grants from the NSF (DBI 0077378 and IBN-0419695 to M.L.G.; DBI 0420126 to J.R.E.) and NIH (DK 30534 to J.K.). T.P. was supported by a training fellowship from the Dartmouth NIH–National Institute of Environmental Health Science Superfund Basic Research Program Project (P42 ES07373) through the Center for Environmental Health Sciences at Dartmouth. A portion of this work was performed at Beamline X26A, National Synchrotron Light Source, Brookhaven National Laboratory. X26A is supported by the U.S. Department of Energy (DOE)–Geosciences and DOE–Office of Biological and Environmental Research, Environmental Remediation Sciences Division.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1132563/DC1
Materials and Methods
Figs. S1 and S2
Table S1
References
Movie S1

17 July 2006; accepted 6 October 2006
Published online 2 November 2006;
10.1126/science.1132563
Include this information when citing this paper.

A NAC Gene Regulating Senescence Improves Grain Protein, Zinc, and Iron Content in Wheat

Cristobal Uauy,[1]* Assaf Distelfeld,[2]* Tzion Fahima,[2] Ann Blechl,[3] Jorge Dubcovsky[1]†

Enhancing the nutritional value of food crops is a means of improving human nutrition and health. We report here the positional cloning of *Gpc-B1*, a wheat quantitative trait locus associated with increased grain protein, zinc, and iron content. The ancestral wild wheat allele encodes a NAC transcription factor (*NAM-B1*) that accelerates senescence and increases nutrient remobilization from leaves to developing grains, whereas modern wheat varieties carry a nonfunctional *NAM-B1* allele. Reduction in RNA levels of the multiple *NAM* homologs by RNA interference delayed senescence by more than 3 weeks and reduced wheat grain protein, zinc, and iron content by more than 30%.

The World Health Organization estimates that more than 2 billion people have deficiencies in key micronutrients such as Zn and Fe and more than 160 million children under the age of 5 lack adequate protein (*1*), leading to an economic burden for society (*2*). The two major types of wheat, tetraploid wheats [diploid cell (2n) = 28], used for pasta, and hexaploid wheats (2n = 42), used primarily for bread, account for ~20% of all calories consumed worldwide. Annual wheat production is estimated at 620 million tons of grain (*3*), translating into approximately 62 million tons of protein. Increasing grain protein content (GPC) has been hindered by environmental effects, complex genetic systems governing this trait, and a negative correlation with yield (*4*). Less progress has been made in increasing Zn and Fe content, the focal point of the HarvestPlus global initiatives (*5*).

Wild emmer wheat [*Triticum turgidum* ssp. *dicoccoides* (DIC)] is the ancestor of cultivated pasta wheat (*T. turgidum* ssp. *durum*) and a promising source of genetic variation in protein, Zn, and Fe content (*6*, *7*). A quantitative trait locus (QTL) for GPC was mapped on chromosome arm 6BS in a population of recombinant inbred lines derived from the *T. turgidum* ssp. *durum* cultivar Langdon (LDN) and the chromosome substitution line LDN (DIC6B) (*8*). This locus was associated with GPC increases of ~14 g kg^{-1} in both tetraploid and hexaploid lines (*8*–*10*). Olmos *et al*. (*11*) mapped this QTL as a simple Mendelian locus, *Gpc-B1* (Fig. 1A), which was later located within a 0.3-cM interval (*12*). Molecular markers *Xuhw89* and *Xucw71* within this region flank a 245-kb physical contig, including *Gpc-B1* (*13*).

Tetraploid and hexaploid wheat lines carrying this 245-kb DIC segment show delayed senescence and increased GPC and grain micronutrients (*14*, *15*). The complete sequencing of this region (DQ871219) revealed five genes

[1]Department of Plant Sciences, University of California, One Shields Avenue, Davis, CA 95616, USA. [2]Institute of Evolution, University of Haifa, Mt. Carmel, Haifa 31905, Israel. [3]United States Department of Agriculture, Agricultural Research Service, Western Regional Research Center, 800 Buchanan St., Albany, CA 94710, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: jdubcovsky@ucdavis.edu

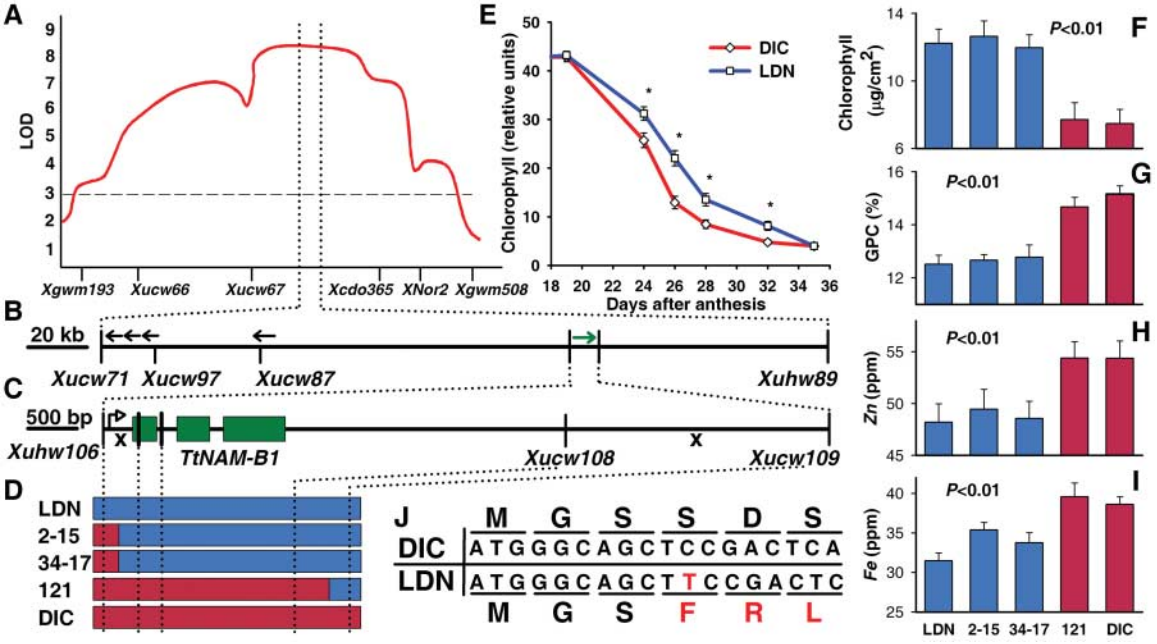

Fig. 1. Map-based cloning of *Gpc-B1*. (**A**) QTL for grain protein on wheat chromosome arm 6BS (*11*). (**B**) Sequenced B-genome physical contig. The position and orientation of five genes are indicated by arrows. (**C**) Fine mapping of *Gpc-B1*. The x's indicate the positions of critical recombination events flanking *Gpc-B1*. Vertical lines represent polymorphism mapped in the critical lines. A single gene with three exons (green rectangles) was annotated within the 7.4-kb region flanked by the closest recombination events. The open arrowhead indicates the transcription initiation site. (**D**) Graphical genotypes of critical recombinant substitution lines used for fine-mapping of *Gpc-B1*. Blue bars represent LDN markers; red bars represent DIC markers. (**E**) Flag-leaf chlorophyll content of recombinant substitution lines segregating for *Gpc-B1* (*14*). Asterisks indicate significant differences ($P < 0.01$). Phenotypes of critical recombinant substitution lines: (**F**) chlorophyll at 20 days after anthesis (DAA), (**G**) grain protein, (**H**) Zn, and (**I**) Fe concentrations. Blue and red bars indicate the presence of the LDN and DIC alleles at *TtNAM-B1*, respectively. (**J**) First 18 nucleotides of DIC and LDN *TtNAM-B1* alleles and their corresponding amino acid translation. The LDN allele carries a 1-bp insertion (red T) that disrupts the reading frame (indicated by red amino acid residues). Error bars represent standard error of the mean.

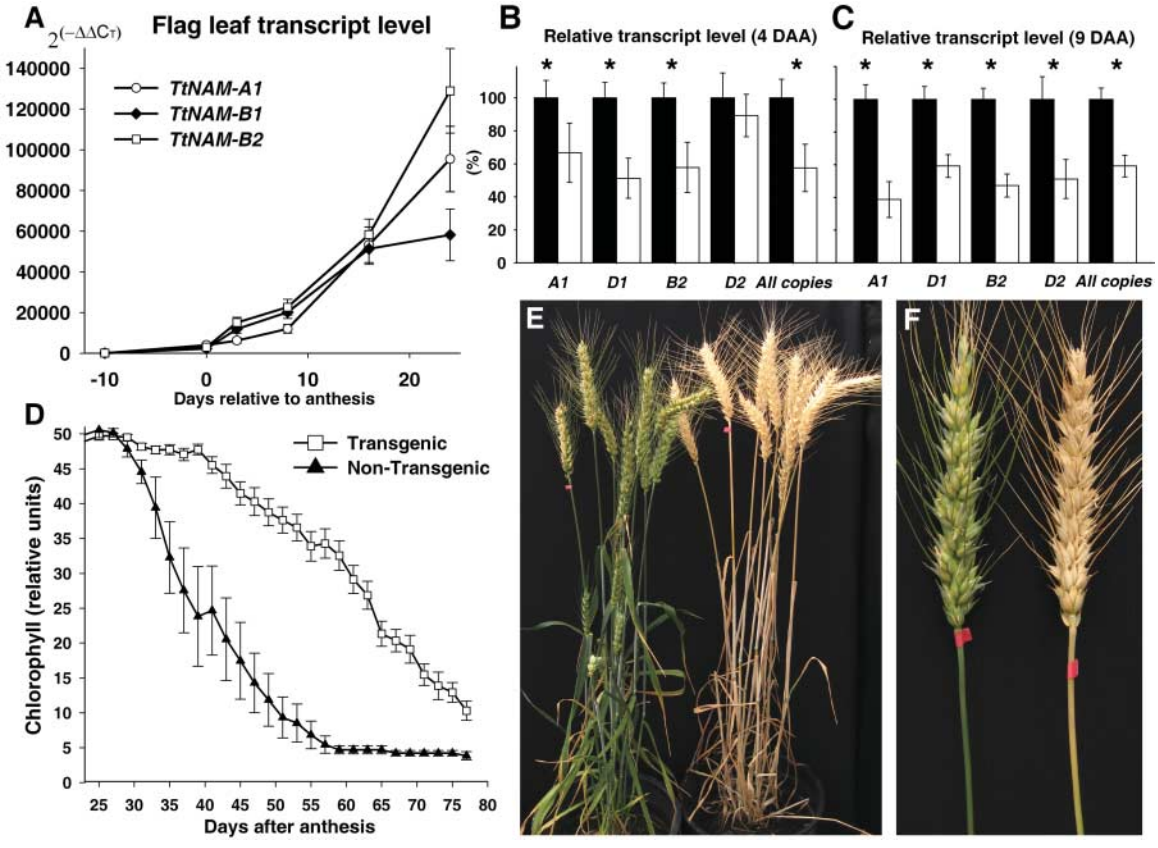

Fig. 2. (**A**) Expression profile of the different *TtNAM* genes relative to *ACTIN* in tetraploid wheat recombinant substitution line 300 carrying a functional *TtNAM-B1* gene. Units are values linearized with the $2^{(-\Delta\Delta CT)}$ method, where CT is the threshold cycle. (**B** and **C**) Relative transcript level of endogenous *TaNAM* genes in T_2 plants (L19-54) segregating for transgenic ($n = 12$, white) and nontransgenic ($n = 11$, black) *TaNAM* RNAi constructs at (B) 4 and (C) 9 days after anthesis. Asterisks indicate significant differences ($P < 0.05$). (**D**) Flag-leaf chlorophyll content profile of transgenic ($n = 22$ T_1 plants) and nontransgenic controls ($n = 10$ T_1 plants). (**E**) Representative transgenic (left) and nontransgenic (right) plants 50 DAA. (**F**) Main spike and peduncles of representative transgenic and nontransgenic plants 50 DAA. Error bars represent standard error of the mean.

(Fig. 1B) (*16*). A high-resolution genetic map, based on approximately 9000 gametes and new molecular markers (table S1), was used to determine the linkage between these genes and the *Gpc-B1* locus. Three recombinant substitution lines with recombination events between markers *Xuhw106* and *Xucw109* delimited a 7.4-kb region (Fig. 1, C and D) (*16*). The recombinant lines carrying this DIC segment senesced on average 4 to 5 days earlier ($P < 0.01$, Fig. 1, E and F) and exhibited a 10% to 15% increase in GPC (Fig. 1G), Zn (Fig. 1H), and Fe (Fig. 1I) concentrations in the grain ($P < 0.01$). Complete linkage of the 7.4-kb region with the different phenotypes suggests that *Gpc-B1* is a single gene with multiple pleiotropic effects.

The annotation of this 7.4-kb region (Fig. 1C) identified a single gene encoding a NAC domain protein, characteristic of the plant-specific family of NAC transcription factors (*17*). NAC genes play important roles in developmental processes, auxin signaling, defense and abiotic stress responses, and leaf senescence (*18*, *19*). Phylogenetic analyses revealed that the closest plant proteins were the rice gene *ONAC010* (NP_911241) and a clade of three *Arabidopsis* proteins including No Apical Meristem (NAM) (figs. S1 and S2). On the basis of these similarities, the gene was designated *NAM-B1* (DQ869673). To indicate the species source, we have added a two-letter prefix (e.g., *Ta* and *Tt* for *T. aestivum* and *T. turgidum* genes, respectively).

Comparison of the parental *TtNAM-B1* sequences revealed a 1-bp substitution within the first intron and a thymine residue insertion at position 11, generating a frame-shift mutation in the LDN allele (DQ869674, Fig. 1J). This frame shift resulted in a predicted protein having no similarity to any GenBank sequence and lacking the NAC domain.

The wild type *TtNAM-B1* allele was found in all 42 wild emmer accessions examined (*T. turgidum* ssp. *dicoccoides*) (table S2) and in 17 of the 19 domesticated emmer accessions (*T. turgidum* ssp. *dicoccum*). However, 57 cultivated durum lines (*T. turgidum* ssp. *durum*) (*20*) (table S3) lack the functional allele, which suggests that the 1-bp frame-shift insertion was fixed during the domestication of durum wheat. The wild-type *TaNAM-B1* allele was also absent from a collection of 34 varieties of hexaploid wheat (*T. aestivum* ssp. *aestivum*), representing different market classes and geographic locations. Twenty-nine of these showed no polymerase chain reaction (PCR) amplification products of the *TaNAM-B1* gene, which suggests that it is deleted, whereas the remaining five lines have the same 1-bp insertion observed in the durum lines (table S4).

In addition to the mutant *TtNAM-B1* copy, the durum wheat genome includes an orthologous copy (*TtNAM-A1)* on chromosome arm 6AS and a paralogous one (*TtNAM-B2*) 91% identical at the DNA level to *TtNAM-B1* on chromosome arm 2BS (*21*) (fig. S3 and table S5). These two copies have no apparent mutations. Comparisons at the protein level of the five domains characteristic of NAC transcription factors (*17*) revealed 98% to 100% protein identity (fig. S2) between barley, wheat, rice, and maize homologs.

Quantitative PCR (*16*) showed transcripts from the three *TtNAM* genes at low levels in flag leaves before anthesis, after which their levels increased significantly toward grain maturity (Fig. 2A). Transcripts were also detected in green spikes and peduncles. The similar transcription profiles and near-identical sequences of *TtNAM-A1*, *B1*, and *B2* suggest that the 4- to 5-day delay in senescence and the 10% to 15% decrease in grain protein, Zn, and Fe content observed in LDN are likely the result of a reduction in the amount of functional protein rather than the complete loss-of-function of a specific gene.

To test this hypothesis, we reduced the transcript levels of all *NAM* copies using RNA interference (RNAi). An RNAi construct (*16*) was transformed into the hexaploid wheat variety Bobwhite, selected for its higher transformation efficiency relative to tetraploid wheat. The RNAi construct targeted the 3′ end of the four *TaNAM* genes found in hexaploid wheat (*TaNAM-A1*, *D1*, *B2*, and *D2*), outside the NAC domain, to avoid interference with other NAC transcription factors (fig. S4 and table S6) (*22*).

We identified two independent transgenic plants (L19-54 and L23-119) with an expected stay-green phenotype. Quantitative PCR analysis of transgenic L19-54 plants showed a significant reduction in the endogenous RNA levels of the different *TaNAM* copies (*22*) at 4 and 9 days after anthesis ($P < 0.05$) (Fig. 2, B and C) compared with control lines. Transgenic plants reached 50% chlorophyll degradation in flag leaves 24 days later than their nontransgenic sibs ($P < 0.001$) (Fig. 2D), and their main spike peduncles turned yellow more than 30 days later than the controls (Fig. 2, E and F).

The presence of the RNAi transgene also had significant effects on grain protein, Zn and Fe concentrations. Transgenic plants showed a reduction of more than 30% in GPC ($P < 0.001$), 36% in Zn ($P < 0.01$), and 38% in Fe ($P < 0.01$) concentration compared with the nontransgenic controls (Table 1). No significant differences were observed in grain size ($P = 0.41$), suggesting that the extra days of grain filling conferred by the reduced *TaNAM* transcript level did not translate into larger grains in our greenhouse experiments (*23*). Similar results were obtained for the second transgenic event, L23-119 (fig. S5 and table S7).

These results suggest that the reduced grain protein, Zn, and Fe concentrations were the result of reduced translocation from leaves, rather than a dilution effect caused by larger grains. This hypothesis was confirmed by analyzing the residual nitrogen, Zn, and Fe content in the flag leaves. We analyzed both transgenic events together (due to greater variability in flag leaves compared with the grains) and confirmed higher levels of N ($P = 0.01$), Zn ($P < 0.01$), and Fe ($P < 0.01$) in the flag leaves of transgenic plants compared with the nontransgenic sister lines (table S8). This supports a more efficient N, Zn, and Fe remobilization in plants with higher levels of functional *TaNAM* transcripts.

These results confirm that a reduction in RNA levels of the *TaNAM* genes is associated with a delay in whole-plant senescence; a decrease in grain protein, Zn, and Fe concentrations; and an increase in residual N, Zn, and Fe in the flag leaf. These multiple pleiotropic effects suggest a central role for the *NAM* genes as transcriptional regulators of multiple processes during leaf senescence, including nutrient remobilization to the developing grain.

The differences observed between the transgenic and nontransgenic plants for these traits were larger than those observed between the LDN and DIC alleles. The RNA interference on all functional *TaNAM* homologs may result in a larger reduction of functional transcripts than the single nonfunctional *TtNAM-B1* allele in tetraploid recombinant lines carrying the LDN allele.

The cloning of *Gpc-B1* provides a direct link between the regulation of senescence and nutrient remobilization and an entry point to characterize the genes regulating these two processes. This may contribute to their more efficient manipulation in crops and translate into food with enhanced nutritional value.

Table 1. Characterization of grain and senescence-related traits of transgenic Bobwhite T_1 plants (event L19-54) segregating for the presence (transgenic, $n = 22$ plants) or absence (nontransgenic, $n = 10$ plants) of the *TaNAM* RNAi construct. TKW, thousand kernel weight; DAA, days after anthesis.

	GPC (%)	Zn (ppm)	Fe (ppm)	TKW (g)	Dry peduncle (DAA)	Dry spike (DAA)
Transgenic	13.27	52.45	37.40	30.23	72.5	53.0
Nontransgenic	19.08	82.50	60.83	31.27	38.4	37.2
Difference	−5.81	−30.09	−23.42	−1.04	+34.1	+15.8
P value	<0.001	<0.01	<0.01	0.41	<0.001	<0.001

References and Notes

1. World Health Organization, *Meeting of Interested Parties: Nutrition* (2001; http://www.who.int/mipfiles/2299/MIP_01_APR_SDE_3.en.pdf).
2. R. M. Welch, R. D. Graham, *J. Exp. Bot.* **55**, 353 (2004).
3. United Nations Food and Agriculture Organization, *Food Outlook* (4, 2005; www.fao.org/documents).
4. N. Simmonds, *J. Sci. Food Agric.* **67**, 309 (1995).

5. www.harvestplus.org/index.html
6. L. Avivi, "High protein content in wild tetraploid *Triticum dicoccoides* Korn," S. Ramanujam, Ed., 5th Int. Wheat Genet. Symp., New Delhi, Indian Soc. Genet. Plant Breed. (1978).
7. I. Cakmak *et al.*, *Soil Sci. Plant Nut.* **50**, 1047 (2004).
8. L. R. Joppa, C. Du, G. E. Hart, G. A. Hareland, *Crop Sci.* **37**, 1586 (1997).
9. A. Mesfin, R. Frohberg, J. Anderson, *Crop Sci.* **39**, 508 (1999).
10. P. W. Chee, E. M. Elias, J. Anderson, S. F. Kianian, *Crop Sci.* **41**, 295 (2001).
11. S. Olmos *et al.*, *Theor. Appl. Genet.* **107**, 1243 (2003).
12. A. Distelfeld *et al.*, *Funct. Integr. Genomics* **4**, 59 (2004).
13. A. Distelfeld, C. Uauy, T. Fahima, J. Dubcovsky, *New Phytol.* **169**, 753 (2006).
14. C. Uauy, J. Brevis, J. Dubcovsky, *J. Exp. Bot.* **57**, 2785 (2006).
15. A. Distelfeld *et al.*, *Physiol. Plant.*, in press.
16. Materials and methods are available as supporting material on *Science* Online.
17. H. Ooka *et al.*, *DNA Res.* **10**, 239 (2003).
18. A. N. Olsen, H. A. Ernst, L. L. Leggio, K. Skriver, *Trends Plant Sci.* **10**, 79 (2005).
19. Y. Guo, S. Gan, *Plant J.* **46**, 601 (2006).
20. M. Maccaferri, M. C. Sanguineti, P. Donini, R. Tuberosa, *Theor. Appl. Genet.* **107**, 783 (2003).
21. The paralogous *TtNAM-A2* copy was not detected in the LDN tetraploid BAC library or by PCR with 2A genome specific primers derived from *T. urartu*, where the *TuNAM-A2* gene is present.
22. The Bobwhite *TaNAM-B1* gene is deleted as determined by PCR with four sets of independent *NAM-B1* specific primers (table S5). Therefore, no expression data are included for *TaNAM-B1* in the transgenic plants.
23. Field experiments including *Gpc-B1* isogenic lines showed a more variable effect of the DIC chromosome region (including *TtNAM-B1*) on grain size (*14*).

24. We thank G. Hart and L. Joppa for the original mapping materials; L. Li, L. Valarikova, and L. Beloborodov for expert technical assistance; R. Thilmony for critical reading of the manuscript; and the National Small Grain Collection, M. Sanguineti, J. Dvorak, and E. Nevo for germplasm used in this study. This research was supported by the National Research Initiative of USDA's Cooperative State Research, Education, and Extension Service, CAP Grant No. 2006-55606-16629, by BARD, the United States–Israel Binational Agricultural Research and Development Fund, Grant No. US-3573-04C, and by a Vaadia-BARD fellowship. GenBank accession numbers are DQ871219 and DQ869672 to DQ869679.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5803/1298/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 to S9
References

9 August 2006; accepted 20 October 2006
10.1126/science.1133649

Evolutionary History of *Salmonella* Typhi

Philippe Roumagnac,[1] François-Xavier Weill,[2] Christiane Dolecek,[3] Stephen Baker,[4] Sylvain Brisse,[2] Nguyen Tran Chinh,[5] Thi Anh Hong Le,[6] Camilo J. Acosta,[7]* Jeremy Farrar,[3] Gordon Dougan,[4] Mark Achtman[1]†

For microbial pathogens, phylogeographic differentiation seems to be relatively common. However, the neutral population structure of *Salmonella enterica* serovar Typhi reflects the continued existence of ubiquitous haplotypes over millennia. In contrast, clinical use of fluoroquinolones has yielded at least 15 independent *gyrA* mutations within a decade and stimulated clonal expansion of haplotype H58 in Asia and Africa. Yet, antibiotic-sensitive strains and haplotypes other than H58 still persist despite selection for antibiotic resistance. Neutral evolution in Typhi appears to reflect the asymptomatic carrier state, and adaptive evolution depends on the rapid transmission of phenotypic changes through acute infections.

Many bacterial taxa can be subdivided into multiple, discrete clonal groupings (clonal complexes, or ecotypes) that have diverged and differentiated as a result of clonal replacement, selective sweeps, periodic selection, and/or population bottlenecks (*1*). Geographic isolation and clonal replacement can also result in phylogeographic differences between bacterial pathogens from different parts of the world (*2*), even within young, genetically monomorphic pathogens (*3*) (supporting online material text) such as *Mycobacterium tuberculosis* (*4*) and *Yersinia pestis* (*5*). Typhi is a genetically monomorphic (*6*), human-restricted bacterial pathogen that causes 21 million cases of typhoid fever and 200,000 deaths per year, predominantly in southern Asia, Africa, and South America (*7*). Typhi also enters a carrier state in rare individuals [such as Mortimer's example of "Mr. N the milker" (*8*)], who can shed high levels of these bacteria for decades in the absence of clinical symptoms. Genome sequences are available from strains CT18 (*9*) and Ty2 (*10*), but the global diversity, population genetic structure, and evolutionary history of Typhi were poorly understood. It has been speculated that Typhi evolved in Indonesia, which is the exclusive source of isolates with the z66 flagellar antigen (*11*).

We investigated the evolutionary history and population genetic structure of Typhi by mutation discovery (*12*) within 200 gene fragments (~500 base pairs each) from a globally representative strain collection of 105 strains. The 200 genes included 121 housekeeping genes; 50 genes encoding cell surface structures, regulation, and pathogenicity; and 29 pseudogenes. Size variation of a poly-T_{6-7} homopolymeric stretch within one gene fragment was inconsistent with other phylogenetic patterns (homoplasies) and this fragment was excluded from further analysis. The other 199

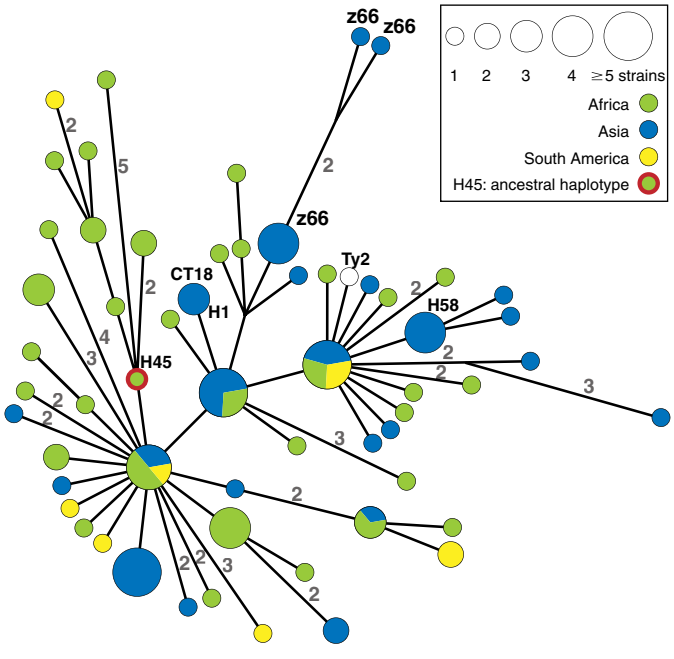

Fig. 1. Minimal spanning tree of 105 global isolates based on sequence polymorphisms in 199 gene fragments (88,739 base pairs). The tree shows 59 haplotypes (nodes) based on 88 BiPs, the continental sources of which are indicated by colors within pie charts. The numbers along some edges indicate the number of BiPs that separate the nodes that they connect; unlabeled edges reflect single BiPs. The genomes of the CT18 and Ty2 strains have been sequenced (GenBank accession codes AL513382 and AE014613, respectively). z66 refers to a flagellar variant that is common in Indonesia (*11*).

[1]Max-Planck-Institut für Infektionsbiologie, Department of Molecular Biology, Charitéplatz 1, 10117 Berlin, Germany. [2]Institut Pasteur, Unité Biodiversité des Bactéries Pathogènes Emergentes, 28 rue du Docteur Roux, 75724 Paris cedex 15, France. [3]Oxford University Clinical Research Unit, Hospital for Tropical Diseases, 190 Ben Ham Tu, District 5, Ho Chi Minh City, Vietnam. [4]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [5]Hospital for Tropical Diseases, 190 Ben Ham Tu, District 5, Ho Chi Minh City, Vietnam. [6]Institut National d'Hygiène et d'Épidémiologie, Hanoi 1000, Vietnam. [7]International Vaccine Institute (IVI), Kwanak Post Office Box 14, Seoul 151-600, Korea.

*Present address: GlaxoSmithKline, King of Prussia, PA 19406–2772, USA.
†To whom correspondence should be addressed. E-mail: achtman@mpiib-berlin.mpg.de